Proceedings of 2013 IAHR Congress

A stochastic Weather Generator for daily climate variable analysis

Lucia Marotta
Phd student, School of Engineering, University of Basilicata, Italy Email: lucia.marotta@email.it

Vito Telesca
Professor, School of Engineering, University of Basilicata, Italy Email: vitotelesca64@gmail.com

ABSTRACT: The objective of this study is to analyse the capability of a Weather Generator based on a multivariate quasi-stationary and weakly depending stochastic process as a tool to take future decisions under the impact of a climate change. A Weather Generator, WG, is a statistical model to generate daily sequences of weather variables, such as precipitation, maximum and minimum temperatures and humidity. Among the different WGs available, there are those built in a two-step process: a first-order Markov chain to generate daily precipitation occurrence, and an exponential distribution to assign daily non-zero precipitation amounts up to a given threshold. ClimGen is widely used as a WG that belongs to this type. At this time, the purpose is to analyse if the most important hypothesis of ClimGen is reliable for the Mediterranean Region, or rather in the following expression $\overline{p}(w/d) = a \cdot \overline{fwet}$ the linear coefficient is constant and equal to $a = 0.75$. The parameter implemented in ClimGen is investigated for ALSIA and ECA stations close to the Mediterranean, particularly for daily precipitation series between 1959-2012. The results show the linear coefficient is not constant and it cannot be assumed as an average value for the analysed dataset because there is no correlation between the output data. The approaches implemented in ClimGen are rough ($a = 1.00$). The methodology has been tested at Policoro station (Basilicata Region, Southern Italy) for which a "*new stochastic model*", that suits climate features including variability in frequency of wet days in a month, has been proposed to generate daily precipitation amounts.

KEY WORDS: Weather Generator, Markov Chain, Climate Change, Wet and Dry Spells, Precipitation Amount.

1. INTRODUCTION

A Weather Generator -WG- is a statistical model to generate realistic daily sequences of weather variables, such as precipitation, maximum and minimum temperatures, humidity, etc. This data is often referred to as synthetic data. Therefore, a WG is not a downscaling technique, it is a model mainly based on statistical hypothesis to obtain synthetic time series of weather data that may be considered as a potential manifestation of the climate. [Brocca et al., 2011]

Traditionally, a WG has been used to overcome problems related to [Johnson et al., 1996; Castellvì et al., 2003]:

- Homogeneous series on a daily basis is not available
- Series on a daily basis is sparse, truncated or difficult to obtain (i.e., expensive, not readily available, etc.)
- Gridding daily weather data for spatial analysis (e.g. of risk)
- Assessing potential impacts when a climate change is presumed in the past and in the future.

Depending on the task to be assessed and the input available, a WG may then be selected. However, it may be recommended to test a number of WGs to check which is more suitable. Different tasks may be considered for selection such as: *data collection, spatial parametrization, model testing and climate scenarios.* [Brocca et al., 2011]
for climate regionalization.

Among the different WGs available there are those built in a two – step process [Hutchinson 1987]. The WG assumes precipitation as primary climate variable that affects the rest of the variables. Thus, the first step consists of modelling daily precipitation. The second step consists of modelling other variables of interest which are affected by precipitation occurrence. We want to obtain a synthetic series of daily maximum and minimum temperatures, solar radiation, humidity and wind speed. In order to implement this, input parameters are usually required on a monthly basis which allows the seasonal variability to be captured and inter-dependency between variables to occur. Daily precipitation is often generated by a two-step process. The first step consists of the generation of wet and dry days. A wet day may be defined by the user, though in general a wet day is considered to have a threshold of rain given by the gauge measurement error, such as 0.25 mm. Markov chains are proven suitable to generate wet and dry days [Richardson and Wright, 1984]. The next step consists of assigning a given amount of precipitation to a wet day. This can be done randomly sampling an appropriate precipitation distribution function, such as a Gamma, Normal, Log-normal, Weibull, etc distribution.

2. THE PROPOSED MODEL FOR THE GENERATION OF PRECIPITATION

The generation of precipitation requires a range of models whose combination and configuration depend on the process and temporal and spatial scales involved. Based on the physical processes involved, three general types of models can be classified [Cox and Isham, 1994]:

a. empirical statistical models, based on stochastic models that are calibrated from actual data. These reproduce annual, monthly and daily precipitation data resembling actual data values.
b. models of dynamic meteorology that incorporate complex non-linear partial differential equations representing different physical processes and that are used for weather forecasting.
c. intermediate stochastic models that incorporate a limited number of parameters determined from actual data collected at short time intervals (for example hourly data) and which are used to represent complicated physical phenomena associated with storm precipitation, such as rain cells, rain bands and cell clusters.

Empirical statistical models for generating daily precipitation data at a given site can broadly be classified into four groups: two-part models, transition probability matrix models, resampling models and ARMA time series models. [Srikanthan & McMahon, 2001].

This study focuses on empirical statistical models [Marotta et al., 2013] which usually represent daily weather sequences, particularly on a type of Two-part model for generating daily precipitation at a specific site. The models reproduce annual, monthly and daily precipitation data resembling actual data values, but, an outstanding problem associated with their use is that they are single-location, or point-process, models. Therefore using these methods for simultaneous simulation of weather sequences at multiple points, for example to evaluate regional hydrological or agricultural behavior, the quite strong spatial correlations in daily weather data must be considered [Castellvì et al., 2003].

Two-part model for daily precipitation consists of two basic steps: first, a model for generating wet and dry events (rainy and non-rainy days); and second, a model for assigning a precipitation amount to a wet day.

Step I: precipitation occurrence

A first-order two state Markov chain is used to stochastically generate dry and wet days, it usually captures the distribution of wet spells as well as higher order models [Racsko et al., 1991; Wilks, 1999].

The generation of precipitation is based on two assumptions. One is that the rain condition on day *i* is related to the rain condition on day *i-1*, and the other is that the amount of rain on rainy days is described by a suitable distribution function.

The first assumption describes a type of model called a Markov chain. Defining $p(w/w)$ as the probability of a wet day on day i given a wet day on day i-1, and $p(w/d)$ as the probability of a wet day on day i given a dry day on day i-1, then the two complementary transition probabilities are:

$$p(d/w) = 1 - p(w/w) \tag{1}$$

that is the probability of a dry day given a wet day on day i-1 and

$$p(d/d) = 1 - p(w/d) \tag{2}$$

that is the probability of a dry day given a dry day on day i-1.

These transition probabilities are calculated for each month at each location of interest. Daily values of these probabilities are interpolated using spline functions. If we know the state of today's weather (wet or dry), we immediately know the probability of a wet day tomorrow. The WG determines whether a particular day is wet or dry by subtracting p(w/w) or p(w/d) from a random number between 0 and 1. If the result is greater than zero, the generator assumes no rain on that day. If it is less than or equal to zero, rain is assumed to have occurred, and the amount of rain is determined using a distribution function for rain amounts on wet days. Quadratic spline functions are used for daily interpolation of monthly probabilities of a wet day given a previous wet day and a wet day given a previous dry day.

As transitional probabilities are conditional, the following expression holds:

$$fwet = p(w/d)\cdot(1-fwet)+p(w/w)\cdot fwet \tag{3}$$

The two transitional probabilities are estimated for each available data source as follows: since the monthly occurrence of precipitation is available the monthly frequency of wet days fwet , can be determined and the transitional probability of a wet day after a dry one p(w/d) for each month is estimated according to the following empirical expression:

$$\begin{aligned} p(w/d) &= 0 && \text{if} \quad fwet = 0 \\ p(w/d) &= a_2\cdot p(w)+a_1 && \text{if} \quad fwet > 0 \end{aligned} \tag{4}$$

where a_1 and a_2 are two-specific coefficients, given by linear regression.
fwet, p(w/d) and p(w/w) on monthly basis are evaluated for a lot of simulated years, but, it’s more important the simulation of spells on a daily basis using the following relations:

$$A = RND\ (0;1) - p(w/d) \tag{5}$$

$$A = RND\ (0;1) - p(w/w) \tag{6}$$

in which RND(0;1) is a random number between 0 and 1.

The first relation is used if the previous day is dry, instead the second is used if the previous day is wet. Then, if A <= 0 the day is wet, instead if A > 0 the day is dry.

Step II: precipitation amount assigned to a wet day.

Nonzero precipitation amounts are simulated here using the exponential distribution. The Gamma and Weibull precipitation distribution functions are selected because their site-specific shape can be estimated from the expected amount of wet day precipitation per month as, respectively, shown in Geng et al. (1986) and Selker and Haith (1990). The model’s distribution function therefore varies from month to month. In this study, the amount of precipitation is assumed to follow a Weibull distribution. According to Rodriguez (1977) the two-parameter Weibull precipitation distribution may be converted to a single parameter distribution using the following expression:

$$W(x,\zeta) = 1-\exp\left[-\left(\Gamma(1+1/\zeta)\frac{x}{\mu}\right)^{\zeta}\right] \tag{7}$$

where x is the daily amount of precipitation, $\mu = \frac{P}{k}$ is the expected monthly amount of wet day precipitation, k is the number of wet days in a month, and ζ is adimensionless parameter related with the coefficient of variation (CV) in the following way [Castellvì et al, 2003]:

$$1+CV^2 = \frac{\Gamma(1+2/\zeta)}{\Gamma(1+1/\zeta)} \tag{8}$$

2.1 Hypotheses of the proposed model

The method is based on different Hypotheses, as follows:

Hypothesis 1. The probability to have a wet day, fwet, is normally distributed.

The normal distribution is a location-scale family. Therefore, it is possible to relate all normal random variables to the standard normal. For example, if X is normal with mean μ and variance σ^2, then

$$Z = \frac{X - \mu}{\sigma} \tag{9}$$

as mean zero and unit variance, that is Z has the standard normal distribution. Conversely, having a standard normal random variable Z we can always construct another normal random variable with specific mean μ and variance σ^2:

$$X = \sigma \cdot Z + \mu \tag{10}$$

therefore:

$$Z = \frac{fwet - \bar{fwet}}{\sigma_{fwet}} = RND \tag{11}$$

where RND is a random number defined in the following way

$$RND = RND(-1;1) \cdot \frac{\bar{fwet}}{3 \cdot \sigma_{fwet}} \tag{12}$$

in which RND(-1;1) is a random number between -1 and 1.

In this way, $\bar{fwet}$ is known as a mean for every month and so monthly basis frequency of wet days is given in the following way:

$$fwet = \bar{fwet} + RND_1 \cdot \sigma_{fwet} \tag{13}$$

Hypothesis 2.

Since the monthly occurrence of precipitation is available, the monthly frequency of wet days, fwet can be determined and the transitional probability of a wet day after a dry one p(w/d) for each month is estimated according to the following empirical expression:

$$p(w/d) = a_2 \cdot p(w) + a_1 \qquad \text{if} \quad fwet > 0 \tag{14}$$

where a_1 and a_2 are two-specific coefficients, given by a linear regression.

If the linear regression between the monthly mean to have a wet day for every period and the monthly mean probability of a wet day after a dry one is like the following

$$\bar{p}(w/d) = a \cdot \bar{fwet} \qquad (a = 1) \tag{15}$$

the probability of a wet day after a dry one, on a monthly basis, can be written in the following way, according to the Eq. (13)

$$p(w/d) = \bar{fwet} + RND_2 \cdot \sigma_{p(w/d)} \tag{16}$$

Then, from the following expression, the monthly probability of a wet day after a wet one can be evaluated

$$fwet = p(w/d)\cdot(1 - fwet) + p(w/w)\cdot fwet \quad (17)$$

fwet, p(w/d), p(w/w) are considered constant within each month because the precise change from day to day is unknown.

Hypothesis 3.

The monthly mean and Standard deviation of frequency to have a wet day is equal for every period (actual and future time):

$$\bar{fwet}(period\ 1) = \bar{fwet}(period\ 2) \quad (18)$$

$$\sigma_{fwet}(period\ 1) = \sigma_{fwet}(period\ 2) \quad (19)$$

The monthly mean and Standard deviation of the probability of a wet day after a dry one is equal for every period (actual and future time)

$$\bar{p}(w/d)\ period\ 1 = \bar{p}(w/d)\ period\ 2 \quad (20)$$

$$\sigma_{p(w/d)}\ period\ 1 = \sigma_{p(w/d)}\ period\ 2 \quad (21)$$

Hypothesis 4.

To evaluate the daily amount of precipitation the followong hypothesis is set:
the coefficient of variation (CV) considered for each month in the actual period is the same for each month in the simulated period.

$$CV\_observed\_period = CV\_simulated\_period \quad (22)$$

3. MATHERIAL AND METHOD

3.1 The database

In this study, a dataset of daily resolution climatic time series is used which has been compiled for the European Climate Assessment (ECA) which 1951-2012. This ECA dataset comprises 199 series of minimum, maximum and/or daily mean temperatures and 195 series of daily precipitation amounts observed at meteorological stations in Europe and the Middle East. In the ECA project, the temperature and precipitation climate is analysed for WMO Region VI (Europe and Middle East: Lebanon, Syria, Jordan and Israel), putting particular emphasis on changes in daily extremes.

Furthermore, a rainfall dataset available for agro-meteorological stations throughout the Basilicata Country is used thanks to the collaboration between ALSIA (Lucanian agency for agricultural development and innovation). For more details about ALSIA station to see [Copertino et al., 2012]. Particularly, only Policoro and Metaponto stations are available for Climatological analyses for the required minimum series length [WMO]. Figure 1 shows the geographical distribution of all stations analysed in this study.

3.2 Homogeneity analysis

Climatic time series typically exhibit spurious (non-climatic) jumps and/or gradual shifts due to changes in station location, environment, instrumentation or observing practices. In many daily resolution climatic time series, there is also a number of missing observation days. Because the degree of inhomogeneity and incompleteness of a daily resolution series determine different Climatic features, data quality control is an ongoing activity in the ECA project.

In the December 2001 version of the ECA dataset, the daily series were subjected to a basic quality control procedure only. Every time series is checked for the occurrence of miscoding, like: precipitation <0 mm; minimum temperature > maximum temperature; non-existent dates; and erroneous outliers. Although the series have usually undergone routine quality control procedures by the supplying institutes, our additional checks identified a number of days with non-correctable mistakes. Such days are assigned ‘missing values’ in the ECA dataset. Currently, statistical homogeneity tests are being applied to the ECA series. The results of these tests will be included in the updated version of the ECA dataset.

Figure 1 Geographical distribution of stations with daily precipitation series.

In any long time series, changes in routine observation practices may have introduced inhomogeneities of non-climatic origin that severely affect the extremes. Wijngaard et al. (2003) statistically tested the daily ECA series of surface air temperature and precipitation with respect to homogeneity. Their methodology has been implemented in ECA&D. A two-step approach is followed. First, four homogeneity tests are applied to evaluate the daily series using the testing variables: the annual mean of the diurnal temperature range DTR (= maximum temperature – minimum temperature); the annual mean of the absolute day-to-day differences of the diurnal temperature range vDTR; the annual wet day count RR1 (threshold 1 mm); the annual number of snow days (SD >= 1cm) SD1; the annual mean of sea-level pressure PP; the annual sum of sunshine duration SS; the annual mean of relative humidity RH and the annual mean of cloud cover CC. The use of derived annual variables avoids auto correlation problems with testing daily series. Second, the test results are condensed for each series into three classes: ‘useful–doubtful–suspect’. Only the ‘useful’ and ‘doubtful’ series can be analysed.

The four homogeneity tests are:

1. Standard Normal Homogeneity Test [SNH, Alexandersson (1986)]
2. Buishand Range test [BHR, Buishand (1982)]
3. Pettitt test [PET, Pettitt (1979)]
4. Von Neumann Ratio test [VON, von Neumann (1941)]

All four tests suppose under the null hypothesis that in the series of a testing variable, the values are independent with the same distribution. Under the alternative hypothesis the SNH, BHR and PET tests assume that a step-wise shift in the mean (a break) is present. These three tests are capable to locate the year where a break is likely. The fourth test (VON) assumes under the alternative hypothesis that the series is not randomly distributed. This test does not give information on the year of the break.[Project team ECA&D, ATBD(2012)].

Whereas, another Homogeneity test (“Test of the runs” [Castellvì F. & Castillo F., 2001]) is carried out for the Basilicata Region series (number of wet days & precipitation amount).

3.3 Application of the models

The two-part model for rainfall simulation consists of a two-state, first-order Markov chain and a two-parameter Weibull probability function. This study focuses on the simplest Markov chain model for rainfall occurrence which includes parameters of two transitional probabilities: from a wet day to a wet day p(w/w) and from a dry day to a wet day p(w/d). Unfortunately, this method requires that many years of daily weather records be available for estimating the model parameters. When these model parameters are evaluated over time and at different places, however, certain general characteristics are revealed.

For example, the transitional probability of a wet day followed by a wet day tends to be greater but parallel to the transitional probability of a dry day followed by a wet day. This phenomenon leads to a strong linear relationship between the *transitional probabilities* and the *fraction of wet days per month*, particularly between p(w/d) and fractions of wet days. Richardson and Wright (1984) have computed the monthly transitional probabilities based on 20 years of data for each of 31 locations in the United States. Their results provide additional information for examining the generality of the linear relationship found. Five locations with different environments were chosen in the United States from Richardson and Wright's report for this purpose: Columbia, Missouri; Boise, Idaho; Miami, Florida; Phoenix, Arizona; and Boston, Massachusetts; Los Banos, Philippines; Wageningen, The Netherlands.

Simple linear regression analysis was performed for each location separately and for the combined data [Geng et al. 1986]. Furthermore, this relationship appears to be independent from each location. The combined regression line, with a zero intercept and slope 0.75, explains 96.5% of the total variation that existed among the transitional probabilities across time and space. The high correlation between p(w/d) and the fractions of the wet days in a wide range of environments leads us to propose the following simple equation:

$$\bar{p}(w/d) = 0.75 \cdot \bar{f}wet \tag{23}$$

At this time, the purpose is to analyse if the basic assumptions of ClimGen [Castellvì et al. 2003] is reliable for the Mediterranean Region. The parameter implemented in ClimGen is investigated for ALSIA and ECA stations close to the Mediterranean area particularly for daily precipitation series between 1959-2012.

The approaches implemented in ClimGen are rough for the Mediterranean Region. The methodology has been tested at Policoro station (Basilicata Region, Southern Italy) for which a "*new stochastic model",* that suits climate features including variability in frequency of wet days in a month, has been proposed to generate daily precipitation amounts.

3.3.1 Application of proposed model and ClimGen

The aim in designing WGs is to produce synthetic weather data which are statistically similar to the observed ones. In this study, they are implemented for Policoro series and statistical tests, comparing the simulated and observed data, are carried out to evaluate their capability in reproducing montlhly climatic patterns which are maybe required for operational purposes in engineering.

New stochastic model - Step I: precipitation occurrence:

Initially, $fwet$, $p(w/d)$, $p(w/w)$ are generated on a monthly basis for 30 years, using Hypothesis 1 and Hypothesis 2. The most important input of the model is $\bar{f}wet$ for each month and the assumptions that the period between 1959/1978, is used as the present, whereas the period around 1993/2012, is used as the future. Then daily sequences of dry and wet days are generated [Eq.(5),Eq.(6)]. At the end, the model is tested looking at the distribution function through the Kolmogorov-Smirnov test between:

- daily data within each month for the observed (20 years) and simulated (30years) probability to find a wet day;
- daily data within each month for the observed and simulated probability to find a wet and dry day;
- daily data within each month for the observed and simulated probability to find two consecutive wet days;
- observed and simulated consecutive sequences of different dry spells in each month;
- observed and simulated consecutive sequences of different dry spells in continuous years;
- observed and simulated consecutive sequences of different wet spells in continuous years.

New stochastic model - Step II: precipitation amount assigned to a wet day:

At first, precipitation amounts are generated for all the wet days (30years) using the Weibull distribution [Eq.(20)] evaluating the expected monthly amount of wet day precipitation μ and the adimensionless parameter ς directly related with the coefficient of variation (CV) for each month for the period around 1960 and around 2012 also using Hypothesis 4. The input for this step is monthly precipitation amount. At the end, the model is tested looking at the mean through the Student's t-test. Data in vectors x and y are independent random samples from normal distributions with equal means and equal but unknown variances, against the alternative that the means are not equal.

ClimGen:

Initially using the 'compact software', precipitation amounts are generated on a daily basis for 30 years and then all climatic probabilities are calculated. Finally, the model is tested looking at the distribution function using the Kolmogorov-Smirnov test for the same features tested for the "*new stochastic model"* and the precipitation amounts are tested looking at the mean using the Student's t-test.

4. RESULTS AND DISCUSSION

4.1 Homogeneity analysis results

The performance obtained from homogeneity test is the following:

Considering the ECA dataset the test results are condensed for each series into three classes: useful–doubtful–suspect. Only the 'useful' and 'doubtful' series are analysed. The quality control results for Policoro and Metaponto stations demonstrate that the series are always homogenous (5%) for number of wet days and precipitation amount in each month.

4.2 Hypothesis of model results

The results about reliability of basic assumptions of ClimGen for the Mediterranean Region demonstrate that a strong linear relationship exists between p(w/d) and fractions of wet days but the linear coefficient cannot be assumed equal to combined value (a=0.75) found by Geng et al. (1986). Simple linear regression analysis is performed for each ALSIA and ECA daily precipitation series close to the Mediterranean Region. Good results are presented in Table 1 in which there are stations with a coherent R^2 [Eq. (15)] and Figure 2 in which the linear relationship appears to be independent from each location, indeed, there is no correlation for the output data.

Table 1 Regression of the transitional probabilities of a dry day to a wet day on the fractions of monthly wet days

STATION	COUNTRY	ID.	a (1959/1978)	R^2	a (1993/2012)	R^2
POLICORO	ITALY	PO1	0.985	0.932	0.941	0.979
METAPONTO	ITALY	PAN	0.826	0.911	0.920	0.908
BRINDISI	ITALY	174	0.579	0.863	0.665	0.899
CAGLIARI	ITALY	175	0.401	0.731	0.586	0.875
BOLOGNA	ITALY	169	0.442	0.816	0.587	0.816
HAR KENAAN	ISRAEL	130	0.603	0.905	0.594	0.840
LIMASSOL	CYPRUS	24	0.645	0.912	0.771	0.933
LARNACA	CYPRUS	23	0.865	0.943	0.799	0.919
PERPIGNAN	FRANCE	36	1.000	0.942	0.928	0.962
SETE	FRANCE	797	0.682	0.890	0.719	0.938
METHONI	GREECE	63	0.549	0.857	0.624	0.866
HERAKLION	GREECE	61	0.785	0.970	0.710	0.942
ZARAGOZA AEROPUERTO	SPAIN	238	0.822	0.934	0.698	0.887
ALBACETE LOS LLANOS	SPAIN	336	0.803	0.945	0.829	0.975
ALICANTE	SPAIN	412	1.010	0.980	0.974	0.900
ALMERIA/AEROPUERTO	SPAIN	3908	1.030	0.923	1.030	0.941
BARCELONA	SPAIN	335	0.968	0.846	1.030	0.898
CASTELLON DE LA PLANA	SPAIN	3925	0.806	0.783	0.861	0.888
GRANADA	SPAIN	417	0.737	0.921	0.732	0.915
IBIZA/ESCODOLA	SPAIN	3916	0.863	0.930	0.861	0.961
MALAGA AEROPUERTO	SPAIN	231	1.000	0.983	0.895	0.965
MURCIA	SPAIN	421	1.080	0.945	0.928	0.962
MURCIA/SAN JAVIER	SPAIN	1404	1.090	0.966	1.040	0.964
REUS/AEROPUERTO	SPAIN	1401	0.557	0.742	0.667	0.734
VALENCIA	SPAIN	237	0.764	0.774	0.681	0.782

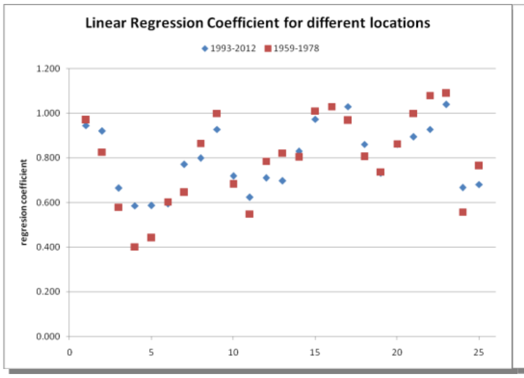

Figure 2 Linear Regression Coefficient.

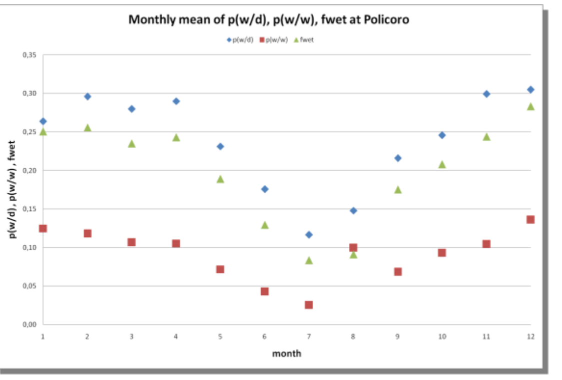

Figure 3 Monthly means of climatic transitional probabilities.

In order to validate the approaches implemented in ClimGen which are rough for the Mediterranean Region, ClimGen is tested for Policoro series. Considering the Hypothesis 0., the basic assumption of Climgen [Castellvì et al. 2003] and the calculated value of the regression coefficient equal to $\overline{a} = 1$, from the theory the following expression can be written

$$\overline{p(w/w)} = \overline{p(w/d)} = \overline{f}\, wet \tag{24}$$

but the following graph [Figure 3] shows that the previous expression [Eq. (24)] is not true for the given location because the magnitude of the different variables are different. Hence, the most important assumption of ClimGen [Eq. (23)] is not valid in Policoro.

Moreover, a “new stochastic model” has been proposed to generate daily precipitation amounts that suits climate features including variability in frequency of wet days in a month.

The following Figures show that for every period the Eq. (15) is true for Policoro series.

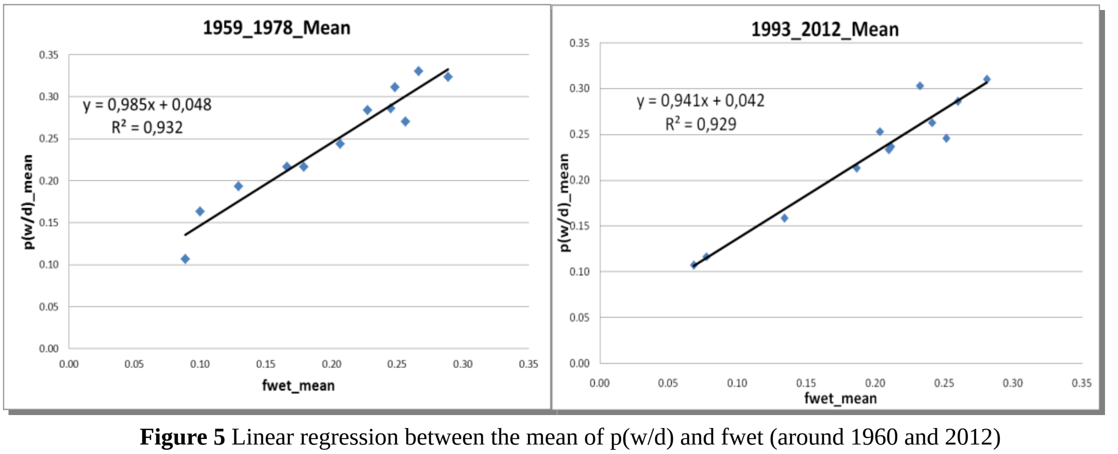

Figure 5 Linear regression between the mean of p(w/d) and fwet (around 1960 and 2012)

4.3 Models results

The following tables show the most important Kolmogorov-Smirnov test results computed to evaluate the capability of the *Step I* for the *New stochastic model* and ClimGen.

Table 2 K-S test results for observed and simulated consecutive sequences of different **dry** spells in each month (*New stochastic model* - around 1960) (yellow means test is true, orange means test is false)

significance level	10% & 5%		Policoro - consecutive dry days - 20 / 30 years – new WG (around 1960)									
type of event	January	February	March	April	May	June	July	August	September	October	November	December
1001												
10001												
100001												
1*5(0)*1												
1*6(0)*1												
1*7(0)*1												
1*8(0)*1												
1*9(0)*1												
1*10(0)*1												
1*11(0)*1												
1*...(0)*1												
1*27(0)*1												
1*28(0)*1												
1*29(0)*1												
1*30(0)*1												
1*31(0)*1												

Table 3 K-S test results for observed and simulated consecutive sequences of different **dry** spells in each month (*ClimGen* - around 1960) (yellow means test is true, orange means test is false)

significance level	10% & 5%		Policoro - consecutive dry days - 20 / 30 years – Climgen (around 1960)									
type of event	January	February	March	April	May	June	July	August	September	October	November	December
1001												
10001												
100001												
1*5(0)*1												
1*6(0)*1												
1*7(0)*1												
1*8(0)*1												
1*9(0)*1												
1*10(0)*1												
1*11(0)*1												
1*...(0)*1												
1*27(0)*1												
1*28(0)*1												
1*29(0)*1												
1*30(0)*1												
1*31(0)*1												

Table 4 K-S test results for observed and simulated consecutive sequences of different **wet** spells in each month (*New stochastic model* - around 1960) (yellow means test is true, orange means test is false)

significance level	10% & 5%		Policoro - consecutive wet days - 20 / 30 years –new WG (around 1960)									
type of event	january	february	march	april	may	june	july	august	september	october	november	december
0110												
01110												
011110												
0*5(1)*0												
0*6(1)*0												
0*7(1)*0												

Table 5 K-S test results for observed and simulated consecutive sequences of different **wet** spells in each month (*ClimGen* - around 1960) (yellow means test is true, orange means test is false)

significance level	10% & 5%		Policoro - consecutive wet days - 20 / 30 years - Climgen (around 1960)									
type of event	january	february	march	april	may	june	july	august	september	october	november	december
0110												
01110												
011110												
0*5(0)*0												
0*6(0)*0												
0*7(0)*0												

The following tables show Student's test results computed to evaluate the capability of the *Step II* for the *New stochastic model* and ClimGen.

Table 6 Student's test results for monthly amount of precipitation (*New stochastic model* - around 1960/2012) (yellow means test is true, orange means test is false)

Policoro- t-test - new WG	around 1960		around 2011	
significance level	5%	10%	5%	10%
number of years	20 / 30		20 / 30	
January				
February				
March				
April				
May				
June				
July				
August				
September				
October				
November				
December				

Table 7 Student's test results for monthly amount of precipitation (*ClimGen* - around 1960/2012) (yellow means test is true, orange means test is false)

Policoro -t-test-Climgen	around 1960		around 2011	
significance level	5%	10%	5%	10%
number of years	20 / 30		20 / 30	
January				
February				
March				
April				
May				
June				
July				
August				
September				
October				
November				
December				

The results obtained suggest that close to Policoro, a first-order Markov chain is capable of capturing dry and wet spells, and that the single-parameter Weibull distribution function is suitable to generate the monthly amounts of precipitation. It has been found that, although the approaches implemented in ClimGen are rough in Policoro, it is capable of explaining dry and wet spells, and monthly precipitation amounts.This is because the probability to have a dry day after a dry day is very high. Therefore, the expression $A = RND - p(w/w)$ to generate rainy events is rarely used. The basis of these WGs are robust to moderate climate changes in precipitation patterns in Policoro because the performance obtained either for the proposed model and ClimGen are good. So they are useful for planning decisions in studies requiring as input dry and wet spells and averages and standard deviations of mean daily precipitation in a month.

5. CONCLUSIONS

The application of WGs may help to take decisions for planning tasks when available weather data is limited. The purpose of this study is the analysis of the capability of a WG built in a two-step process, first-order Markov chain to generate daily precipitation occurrence and an exponential distribution to assign daily non-zero precipitation amounts up to a given threshold.

ClimGen is widely used as WG that belongs to this type. At first, this study focuses on the reliability of basic assumptions of ClimGen [Eq.(1), Eq.(36)] for the Mediterranean Region. Indeed, a strong linear relationship exists between p(w/d) and fractions of wet days but the linear coefficient cannot be assumed equal to the combined value found by Geng et al. (1986). Simple linear regression analysis has been performed for ALSIA and ECA daily precipitation series close to the Mediterranean. Results show that the linear relationship appears to be independent from each location. Infact, there is no correlation between the output data. The assumptions of ClimGen are rough for the Mediterranean Region and so ClimGen has been tested at Policoro station. Considering the Hypothesis 0. of the model. [Eq. (1)] and the calculated value of the regression coefficient equal to ($a = 1$), an error has been found between the ClimGen theory [Eq. (37)] and Policoro climatic features.

Therefore, the "*New stochastic model*" has been proposed to generate daily amounts of precipitation that suits climate features including variability in frequency of wet days in a month.

Statistical tests (*K-S test and Student's t-test*) are then conducted to evaluate the capability in reproducing montlhly patterns which are perhaps required for operational purposes in engineering. The results obtained suggest that close to Policoro, a first-order Markov chain is capable of capturing dry and wet spells, and that the single-parameter Weibull distribution function is suitable to generate the monthly amounts of precipitation. It has been found that, although the approaches implemented in ClimGen are rough in Policoro, it is capable of explaining dry and wet spells, and monthly precipitation amounts.This is because the probability to have a dry day after a dry day is very high. Therefore, the expression
$A = RND - p(w/w)$ to generate rainy events is rarely used. The basis of these WGs are robust to moderate climate changes in precipitation patterns in Policoro because the performance obtained either for the proposed model and ClimGen are good. So they are useful for planning decisions in studies requiring as input dry and wet spells and averages and standard deviations of mean daily precipitation in a month.

Therefore, the models reproduce the main features of precipitation required for agricultural, forestry and civil planning with regard to taking decisions concerning irrigation, potential agricultural productions, and managing the risk of extreme events and disasters due to climate change, such as frequency of drought or flood events. This encourages us to carry out future studies to test the models at other sites with similar climate patterns, such as around the Mediterranean Region. Moreover, this WG can be implemented to obtain any synthetic series of primary weather data, such as air temperature.

ACKNOWLEDGEMENT

The authors would like to thank ALSIA (Lucanian agency on development and innovation in agriculture) and ECA (European Climate Assessment) for providing the dataset.

References

Brocca L., Camici S., Tarpanelli A., Melone F. & Moramarco T.,2011. Climate change effects on floods frequency through a continuous hydrological modelling. 3rd Internacional Meeting on Meteorology andClimatology of the Mediterranean 6-9 June,2011 – Castellaneta Marina.

Castellvì F., Mormeneo I.,Perez P.J., 2003. Generation of daily amounts of precipitation from standard climatic data: a case study for Argentina. Journal of Hydrology 289 (2004) 286–302.

Castellvì F., Castillo F.,2001. Agro-meteorologia, chapther 19, 447-482.

Copertino V.A., Di Pierro M., Scavone G. and Telesca V., 2012. Comparison of algorithms to retrieve Land Surface Temperature from LANDSAT-7 ETM+ IR datain the Basilicata Ionian band. Tethys - Journal of Mediterranean Meteorology & Climatology **9**: 25 – 34. ISSN 1697-1523; eISSN 1139-3394. DOI: 10.3369/tethys.2012.9.03. Edited by ACAM.

IPCC, 2012. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp.

Klein Tank A.M.G. et al., 2002. Daily dataset of 20-th-century surface air temperature and precipitation series for the European Climate Assessment. Int. J. Climatol. 22, 1441-1453.

Geng, S., Frits, W.T., de Vries, P., Supit, I.,1986. A simple method for generating daily rainfall data. Agric. For. Meteorol. 36, 363–376.

Linacre E.,1925. Climate data an resources: a reference and guide. Part I Climate data, chapter n. 4, 109-138.

Marotta L., Castellvì F., Telesca V., 2013. A stochastic Weather Generator for daily precipitation generation: a study case for Policoro - Basilicata Region (IT) 4th International Meeting on Meteorology and Climatology of the Mediterranean-Banyuls (Roussillon, France), February 27 - March 1, 2013

Project team ECA&D, 2012. European Climate Assessment & Dataset (ECA&D).

Richardson C.W., and Wright D.A., 1984. WGEN: A model for generating daily weather variables. U.S. Department of Agriculture, Agricultural Research Service, ARS-8, 83 pp.

Srikanthan R., Mcmahon T.A,2001. Stochastic generation of annual, monthly and daily climate data: A review. Hydrolody and Herat System Sciences 5, 4 (2001) 653-670.